Summary of MODIS Version 2 Requirements for Time-Dependent Ancillary Data

White Paper

By: Barry E. Herchenroder

MODIS Science Data Support Team (SDST)
7501 Forbes Blvd., Suite 103
Seabrook, MD 20706

beh@ltpmail.gsfc.nasa.gov
301-352-2105

Acknowledgment

This white paper is an updated and expanded version of an informal summary of MODIS Version 2 ancillary data requirements written by Dr. Richard P. Cember, the previous Ancillary Data Lead of the MODIS SDST. I greatly appreciate his efforts developing the summary, the material provided within this document, and the numerous discussions about ancillary data that we have had.

Summary of MODIS Version 2 Requirements for Time-Dependent Ancillary Data

White Paper

Table of Contents

Summary of MODIS Version 2 Requirements for Time-Dependent Ancillary Data

White Paper

1. INTRODUCTION

This document was prepared to outline the current state of the ancillary data activity for Moderate Resolution Imaging Spectroradiometer (MODIS). It is meant to provide a clear overview of the current ancillary data situation for Version 2. Before getting into specific details, a few general points are appropriate.

1.1 Ancillary Data Overview

The term “ancillary data” in the MODIS production environment excludes MODIS products themselves. The reason for this exclusion is that MODIS products require no special arrangements whereas non-MODIS products are not automatically available in the production environment and therefore access to them needs to be arranged with the data provider (source). Ancillary data typically are used for two purposes. One is for creation of the MODIS product itself and the other is for validation of the created product. Only those files used in the generation of the product are addressed in this paper. Additionally, some ancillary data files contain static information (such as the Digital Elevation Model [DEM]) and others contain time-dependent information (such as the TOVS twice daily ozone product). These types of ancillary data are further discussed in the following paragraphs.

1.2 Static Ancillary Versus Time-Dependent Ancillary Data Sets

This document concerns time-dependent ancillary data exclusively, except for a brief discussion of static data sets in this paragraph. Static ancillary data sets fall essentially into two categories:

1. Data sets supplied for use with the Earth Observing System (EOS) Data and Information System (EOSDIS) Core System (ECS) Science Data Processing (SDP) Toolkit, such as DEMs, the Zobler soil data set, etc.; and

2. Static data sets or look-up tables supplied by each Science Computing Facility (SCF) for use with their code.

Neither of these requires any ongoing action by the Science Data Support Team (SDST). However, SDST does attempt to track the development of the upcoming revised DEM tool because it is used by multiple MODIS developers.

1.3 General Nature of the SDST Activity for V2 Ancillary Data

The SDST ancillary data activity reflects the overall nature of how time-dependent ancillary data (hereafter, simply "ancillary data") will function in the Distributed Active Archive Center (DAAC) environment.

During actual processing, what the DAAC environment will be providing to a running Product Generation Executive (PGE) are ancillary data files; what Earth Science Data and Information System (ESDIS)/ECS is requesting from SDST is the specification of the ancillary files that the DAAC will need to make available at runtime, including the file names and the servers where the needed files can be found. Thus, the basic unit of ancillary data is the file, not the field (or array) in the file. Science code is assumed to be able to read the files that contain the fields of interest.

Thus, SDST seeks to work with ancillary data needs at the file level. ESDIS/ECS is not concerned with what fields are needed and is not designing any mechanism to extract fields for a given piece of science code.

1.4 Ancillary Data and HDF-EOS

All ancillary data sets will be available in native format (that is, as provided by National Oceanic and Atmospheric Administration (NOAA), or whoever), and will be archived in native format. In addition, for ancillary data files requested by more than one instrument team, the DAAC will reformat that data to Hierarchical Data Format-EOS (HDF-EOS) using reformatting tools written by ESDIS/ECS, and archive the reformatted data in addition to the native data. SCFs are under no obligation to use the reformatted data but they may do so if they wish. Some reformatting tools have already been written for some data sets of NOAA origin; these are available to the science teams if they want to use them for development purposes.

In the future, ESDIS/ECS may support subsetting or other operations on ancillary files below the file level, but this will only be for files that come in, or are reformatted to, HDF-EOS.

2. DATA SOURCES

Time-varying ancillary data has been requested from many sources. The sources identified for V2 so far consist of the following:

NASA Sources

- Data Assimilation Office (DAO)
- Atmospheric Chemistry Group at Goddard Space Flight Center (GSFC)
- Jet Propulsion Laboratory (JPL) NSCAT Processing Group
- JPL Physical Oceanography (JPL-PO) DAAC

NOAA Sources

- National Centers for Environmental Prediction (NCEP), formerly the National Meteorological Center (NMC)
- National Environmental Satellite Data and Information Service (NESDIS)

Others

- U.S. Navy's Fleet Numerical Meteorological and Oceanographic Center (FNMOC)
- Woods Hole Oceanographic Institute (WHOI)
- U. of Miami Ocean Remote Sensing Group
- Other Satellite Sensors (not considered ancillary data but included for completeness):
  - MISR
  - AIRS/AMSU

Below is a short discussion about each data source. It is SDST's understanding that data from each source (except AIRS/AMSU which is a PM-1 instrument) will be available for AM-1 at the GSFC DAAC. Any of the ancillary data sets needed for MODIS Land processing at the EROS Data Center (EDC) DAAC will in all likelihood have to be transferred from GSFC to EDC by ftp. It is not known yet what the details of the transfer process will or whether this process will be transparent to the PGEs that need the data set.

2.1 NASA Sources

2.1.1 Data Assimilation Office

The current plan is for DAO model output data to be provided to ECS, and by implication AM-1 and MODIS, in HDF-EOS. This data, not yet produced operationally, is referred to as the "DAO high time resolution data set." The name is fitting since its output time interval will be significantly smaller than the six hour (at best) time interval of the corresponding NCEP Meteorological (Met) model output.

The initial operational version of the high time resolution data (available in late V2 and the initial AM-1 Post Launch (PL) period) will use a grid spacing of 2.5 degrees longitude by 2.0 degrees latitude. A 34 day sample data set covering the same time period as the MODIS V1 ancillary data (July 29, 1996 to September 1, 1996) will be available shortly, probably by the middle of May 1997.

A file specification document for the DAO data set is currently being written and will be completed shortly, probably by the middle of May 1997.

2.1.2 Atmospheric Chemistry Group at GSFC

This group prepares Near Real-Time (NRT) TOMS Column Ozone Data from the raw TOMS data. TOMS Column Ozone NRT data has been requested by MODIS and will be available at the GSFC DAAC in both native and HDF formats (**NOT** in HDF-EOS!).

2.1.3 JPL NSCAT Processing Group

The only data set identified up to now from this source is possibly the NRT ADEOS NSCAT Overwater Surface Winds. It is not known yet whether this data set will be acquired from the JPL NSCAT Processing Group or from NOAA/NESDIS. At present, there are no plans to offer this data set in HDF-EOS; only the native format will be available at the GSFC DAAC.

2.1.4 JPL-Physical Oceanography DAAC

NRT Pathfinder AVHRR Sea Surface Temperature (SST) data has been requested from the JPL Physical Oceanography DAAC, or JPL-PO DAAC, This data, written in HDF, will become available sometime in early Calendar Year (CY) 1998.

2.2 NOAA Sources

2.2.1 NCEP

NCEP data consists of both model output data and the equivalent of Level 3 gridded observed data. The native format of all the data except Reynold SST at present, is the

World Meteorological Organization (WMO) Grib format (NOAA Office Note 0N388 Edition 1). The native form of Reynolds SST is ASCII.

One data set (TOVS Twice Daily Column Ozone Product - see Section 4.2.2) may have its native format switched from Grib to the WMO BUFR format in the near future because of a change in the NCEP computer platform which creates the data set.

NCEP Grib data sets will also be available in HDF-EOS. However at present, all the Science Teams that have requested NCEP data plan to use it in its native form, not in its HDF-EOS form.

2.2.2 NESDIS

The only data set identified up to now from this source is the NRT ADEOS NSCAT Overwater Surface Winds. It is not clear yet whether this data set will be acquired from NOAA/NESDIS or from the JPL NSCAT Processing Group. At present, there are no plans to offer this data set in HDF-EOS; only the native format will be available at the GSFC DAAC.

2.3 Others

2.3.1 Navy's FNMOC

The only data set identified from the Navy's FNMOC has been the global ocean mixed-layer numerical model output product. The native format of this data set is grib. At present, there are no plans to offer this data set in HDF-EOS; only the native format files (or standard HDF format files if WHOI supplies the data) will be available at the GSFC DAAC.

2.3.2 Woods Hole Oceanographic Institute

The Ocean-Atmosphere Group is a possible alternative to the Navy as a source for oceans mixed layer depth numerical model output product. This group acquires the Navy native mixed layer depth files as they become available, converts them to standard HDF, and places them on a server accessible by ftp. There are no plans to offer this data set in HDF-EOS.

2.3.3 University of Miami Ocean Remote Sensing Group

Buoy Time-and Location Data compiled by the University of Miami Ocean Remote Sensing Group has been requested by Bob Evans as input to his Ocean Color and SST Matchup Database process. The native format of this data is ASCII. There are no plans to offer this data set in HDF-EOS; only the native format will be available at the GSFC DAAC.

2.3.4 Other Satellite Sensors

The following are other satellite sensors that are not considered ancillary data but are included for completeness:

- MISR – Data from this source is desired by some members of the Land Team. Since MISR is an AM-1 instrument, it is not considered ancillary data because it will have no interface to non-EOS data.

- AIRS/AMSU – This is a PM-1 instrument that has been requested for MODIS. Like MISR, it is not considered ancillary data because it will have no interface to non-EOS data.

3. SCIENCE TEAM ANCILLARY DATA NEEDS

The primary V2 ancillary data sources for the Atmosphere and Land Teams will be the DAO high time resolution model output data set. NCEP data will be the primary ancillary data source for the Miami Ocean Team. SST and Mixed Layer Model output will be the primary data source for the Esaias/Turpie Ocean Team.

Below are further details about the MODIS Science Team ancillary data needs.

3.1 Atmosphere Team

SDST's understanding is that the primary ancillary data source for all atmosphere processes will be the DAO high time resolution model output data set. The backup data sets for V2 are not fully defined.

3.2 Ocean Team

SDST's understanding from the Miami team is that their primary ancillary data needs are identical to those of SeaWiFS. These needs will be supported by SeaWIFS ancillary data routines.

Bob Evans of the University of Miami has indicated that the Miami Team would like to use the DAO high time resolution data as an ancillary data backup. Miami also will use climatology (supplied by SeaWiFS) as a backup if NCEP and DAO data are not available.

The Esaias/Turpie (Ocean Primary Productivity Team) ancillary data needs are understood to primarily consist of SST data and Ocean Mixed Layer model output. The team will use climatological data as a backup for the missing ancillary data when one or both of their primary ancillary data source(s) is not available.

3.3 Land Team

The Land Team plans to make only modest use of ancillary data since most of its input data needs are met by output from other MODIS products. Thus for the Land Team, ancillary data is itself usually a backup. It is SDST's understanding that the primary V2 ancillary data sources for the Land Team will be the DAO High Time Resolution Model Output data set. At present most members of the Land Team will be using built-in climatology as their primary ancillary backup.

4. SPECIFIC ANCILLARY DATA SETS

A summary of the V2 time-varying ancillary data needs as SDST understands them is given in the spreadsheet included in Appendix A and in the sections that follow.

If any items in that appendix are incorrect or items are missing, SDST needs to know as quickly as possible so that ECS can be informed and arrangements can be made to acquire the data for V2.

4.1 Access to Ancillary Data Sets

For the present, data sets are fetched by the ESDIS/ECS "larry" server (ftp: larry.gsfc.nasa.gov) at the GSFC DAAC. The data sets are then stored on larry in a sub-directory tree starting with one of the root directories (/pub/ncep_data, pub/data, /pub/adeos, or /pup/eptoms) and made available via an anonymous ftp site.

Due to limited disk space, only one month of data is kept on line. At the end of the month, the data is stored on backup tape. If backed up data is needed, a request for it can be made to Ms. Qiulian Yang (Email: qyang@daac.gsfc.nasa.gov) who is in charge of the larry server at the DAAC. She will write the desired data on a temporary disk and make it available to the requester.

Data sets will **NOT** be accessible directly from the larry server under operational conditions. The manner of accessing data under operational conditions has yet to be completely specified by ESDIS/ECS. Operational access to time-varying ancillary data will be through the ECS Ingest Server. It appears that access will involve use of production rules and the insertion of appropriate file "Universal References" for each time-varying ancillary data file into the Process Control File (PCF) file associated with a given PGE.

4.2 Version 1 Ancillary Data Used in Version 2

Below is a detailed summary of each data set identified so far for V2. YY, MM, DD refers respectively to the year (00 to 99), month (01 to 12), and day (01 to 28, 29, 30, or 31 depending on the month), of the fields that a given file contains.

Note that the HDF-EOS versions of Grib files are not being produced at the GSFC DAAC yet. It is SDST's understanding that these HDF-EOS files will not be produced operationally until late in 1997. It is also not known yet when ECS will have "official" samples of the Grib files in HDF-EOS format available. SDST is pursuing this issue.

4.2.1 2.5 Degree NCEP GDAS Final Run

This analysis comes out every six hours and is being used by SeaWiFS. Therefore, 2.5 degree Global Data Assimilation Model (GDAS) is the Met analysis for the MODIS ocean team. It is distributed by the NCEP "nic" server (ftp: nic.fb4.noaa.gov). The "larry" server (ftp: larry.gsfc.nasa.gov) at the GSFC DAAC fetches the 2.5 degree GDAS files daily from the nic server and stores it. The data is then made available via an anonymous ftp site on larry.

This data set will no longer be produced by NCEP after the middle of March 1997. As a result, SeaWiFS and MODIS have decided to switch from the 2.5 degree GDAS data set to the 1.0 degree GDAS data set as of roughly March 15, 1997. The SeaWiFS ancillary data interfaces are being modified to use the 1 degree GDAS data instead of the 2.5 degree GDAS data. MODIS will use the revised SeaWiFS ancillary data interface and supporting code and files as soon as they become available from the SeaWiFS project.

The fetch information (until the middle of March 1997) for the nic and larry servers is as follows:

nic pathname: /pub/data.XXz/ced1.ganl.tHHz.fnl (XX = 00 has AM analyses; XX = 12 has PM analyses; HH = 00, 06 for XX = 00; HH = 12, 18 for XX = 12.)

larry pathname: /ncep_data/nmc_YYMM/anal_YYMM/fnl.ced1.ganl.YYMMDD.HHz (HH = 00, 06,12,18)

Frequency (files/day) = 4

Native Format = Grib

Note that the larry filename contains the date and the nic filename does not. The files available on the nic servers are updated daily. The larry server keeps the files on disk for the current calendar month.

4.2.2 TOVS Column Ozone Twice-Daily Product

This product is used by SeaWiFS, and therefore by the MODIS ocean team. As with the 2.5 degree GDAS, it is distributed by the nic server, fetched daily (eventually twice daily) by the larry server, and made available via an anonymous ftp site.

nic pathname: /pub/data.XXz/satnss.tovs.tHHz.cur (XX = 00 has AM analyses; XX = 12 has PM analyses; HH = 00 for XX = 00; HH = 12 for XX = 12)

larry pathname for xx=00 file:
/pub/ncep_data/obs_YYMM/tovsYYMM/tovsYYMMDD.HHz (HH = 00, 12)

Frequency (files/day) = 2

File size (MB/file) = 7.61

Volume (MB/day) = 15.22

Native Format = Grib

Formats Ultimately Available for V2: Grib, HDF-EOS

Note again that the larry filename contains the date and the nic filename does not. The files available on the nic server are updated daily. The larry server keep the files on disk for the current calendar month. A twice daily fetch of these products files has been requested from the GSFC DAAC. The 0000 UTC fetch has been implemented and the implementation of the 1200 UTC fetch is expected by the end of May 1997.

The TOVS Column Ozone Twice-Daily Product data set may possibly have its native format changed from Grib to the WMO BUFR format in the near future due to a change in the NCEP computer platform which creates the data set. SDST will track this possible format change and inform the Science Teams (particularly the Ocean Team that will ingest it using a SeaWIFS-like ancillary data interface) when, and if, such a format change takes place.

4.2.3 1 Degree NCEP Aviation Model (AVN) Forecast

This is currently the Met model input for the Wisconsin ancillary preprocessor and therefore for the atmosphere team. This forecast comes out twice daily, is distributed by the NCEP nic server, fetched daily (eventually twice daily) by the larry server, and made available via an anonymous ftp site.

nic pathname: /pub/avn/avn.XXz/gblav.THHZ.PGrbFHH (XX = 00, 12; HH =00, 03, 06,...,72)

larry pathname for xx=00 file:
/pub/ncep_data/anal_YYMM/anal_YYMMDD/gblav.PGrbF00.YYMM.HHz

Frequency (files/day) = 2

File size (MB/file) = 15.9

Volume (MB/day) = 2 X 15.9 = 31.8 (for the analysis fields only–these have XX=00, 12)

Native Format = Grib

Formats Ultimately Available for V2: Grib, HDF-EOS

A twice daily fetch of this forecast by the larry server has been requested from the GSFC DAAC. The 0000 UTC fetch has been implemented and the implementation of the 1200 UTC fetch is expected by the end of May 1997.

4.2.4 TOVS Column Ozone Daily Product

This is currently the ozone input for the Wisconsin ancillary preprocessor, therefore, for the atmosphere team. The product files come out once daily, are distributed by the

NCEP nic server, fetched daily by the larry server, and made available via an anonymous ftp site.

nic pathname: /pub/nws/nmc/cac/htdocs/products/stratosphere/tovsto/YYMMDD.grb

larry pathname: /pub/data/tovsl_YYMM/YYMMDD/grb

Frequency (files/day) = 1

File size (MB/file) = .098

Volume (MB/day) = .098

Native Format = Grib

Formats Ultimately Available for V2: Grib, HDF-EOS

This product is quasi-daily, in that sometimes NCEP misses a day.

4.2.5 Reynolds Weekly Average SST Product

This has been requested both by the Wisconsin and the Miami teams. Distribution is by the NCEP nic server. Each product file covers a week from Sunday to Saturday and is given the date of the Wednesday in the middle of this week. The product files are fetched weekly by the larry server and made available via an anonymous ftp site.

nic pathname: /pub/ocean/clim1/oiupdate/oi.mean.bias.YYMMD'D'
(D'D' = day of the month of the Wednesday of the week to which the data in the file corresponds).

larry pathname: /pub/data/R_SST_Y/oi.mean.bias. YYMMD'D'

Frequency (files/day) = 1/ 7

File size (MB/file) = 0.262

Volume (MB/day) = 0.262/ 7

Native Format = ASCII

Formats Ultimately Available for V2: Native

4.3 New Version 2 Ancillary Data Sets

The poll of the Science Teams that was sent out several months ago over Ed Masuoka's e-mail signature brought to light several additional ancillary needs and desires for Version 2 and thereafter. These are summarized in the following sections.

4.3.1 DAO High Time Resolution Global Analysis Model Output in Support of AM-1

This has been requested by the Justice and Vermote Land Team (at GSFC) and the Running & Glassy (at the University of Montana) Land Team and by the University of

Miami Ocean Team. DAO output will consist of one file per DAO model output time level. The file will be produced in the HDF-EOS format. The output time interval will be significantly smaller than the six hour (at best) time interval of the corresponding NCEP Meteorological (Met) model output. Surface data will be output hourly and Upper Air data will be output at three hour intervals.

The initial operational version of the high time resolution data (available in late V2 and the initial AM-1 PL period) will use a grid spacing of 2.5 degrees longitude by 2.0 degrees latitude. The follow-on version of the high time resolution data (available about 6-9 months after AM-1 launch) will use a grid spacing of 1.0 degree by 1.0 degree.

DAO and ECS have agreed upon the V2 HDF-EOS format for the initial operational version of the DAO high time resolution data set (in particular, the metadata that will be included in the data set). A file specification document for this data set will be completed shortly (expected by the middle of May 1997). When finished, this file specification will be available at the DAO ftp site: hera.gsfc.nasa.gov. A discussion of the variables to be supplied in this data set is included in a handout supplied by Jim Stobie of the DAO at the Algorithm Developers Forum held in October 1996 during the MODIS Science Team Meeting.

A one file sample of the initial operational version of the high time resolution data set is located in the sub-directory /tmp/ravi at the same ftp site. The file name is DAOE054A.hdf. A 34 day sample of the initial operational version of the high time resolution data set will be available shortly (expected by mid-May 1997). The data will cover the same time period as the MODIS V1 ancillary data (*viz*., July 29, 1996 to September 1, 1996). Sample files with the follow-on 1.0 degree by 1.0 degree resolution model output data will probably not be available until early CY 1998.

4.3.2 SSM/I Daily Sea Ice Product

The Special Sensor for Microwave/Imaging (SSM/I) was requested by the University of Wisconsin Atmosphere Team. Distribution is by the NCEP “polar” server (ftp: polar.wwb.noaa.gov) supported by the NCEP Ocean Modeling Branch. Two files are generated for each day–one for northern hemisphere ice and one for ice in the southern hemisphere. The product files are fetched once a day by the larry server and made available via an anonymous ftp site.

polar pathname (northern hemisphere ice): /pub/ice/northpsg.YYMMDD

larry pathname: /pub/data/ssmi_YYMM/northpsg.YYMMDD

Frequency (files/day) = 1

File size (MB/file) = 0.179

Volume (MB/day) = 0.179

Native Format = Grib

Formats Ultimately Available for V2: Grib, HDF-EOS

polar pathname (southern hemisphere ice): /pub/ice/southpsg.YYMMDD

larry pathname: /pub/data/ssmi_YYMM/southpsg.YYMMDD

Frequency (files/day) = 1

File size (MB/file) = 0.123

Volume (MB/day) = 0.123

Native Format = Grib

Formats Ultimately Available for V2: Grib, HDF-EOS

4.3.3 FNMOC Global Mixed-Layer Model Product (U.S. Navy's)

This has been requested by the MOD27Y oceans group (Esaias & Turpie). This data has here-to-fore been distributed by the "Navy NODDS" dial-up data distribution system responsible for disseminating Navy generated numerical model output to civilians and non-DOD U.S. Federal Government Agencies. This dial-up arrangement is not acceptable to ESDIS/ECS. Therefore arrangements to acquire this data by the more conventional (and satisfactory to ESDIS/ECS) ftp approach are being worked out by SDST and ESDIS/ECS.

SDST has discovered that the mixed layer model product is being acquired daily by WHOI, converted to HDF, and placed on a WHOI server accessible by ftp. This information has been given to ESDIS. They are exploring the possibility of acquiring the HDF files from the WHOI server via ftp. Alternately, an arrangement may be made with the NOAA NODDS distribution system to upgrade their capability to allow access of the native mixed layer files by ftp.

Whichever of these two approaches is finally adopted, the files would be accessed by the GSFC DAAC using ftp and placed on the larry server for use by ESDIS/ECS and ultimately by the MOD27Y science code. This data set is only offered in the native Grib formatted files (if files are ultimately accessed from a NODDS server) or HDF formatted files (if files are ultimately accessed from the WHOI server). There are no current plans to offer this data set in HDF-EOS.

Frequency (files/day) = TBD

File size (MB/file) = TBD

Volume (MB/day) = TBD

Native Format = Grib

Format Ultimately Available for V2: Native Grib (if files are ultimately accessed from a NODDS server) or HDF (if files are ultimately accessed from the WHOI server).

4.3.4 Pathfinder AVHRR SST Product

This has been requested for MOD27Y by the oceans group (Esaias & Turpie). This data is generated by Bob Evans, University of Miami oceans group in HDF and then uploaded to the JPL Physical Oceanography (JPL-PO) DAAC. There it becomes available to requesters in several different formats.

At present Miami is generating this product retrospectively and has completed SST calculations through CY 1991. The JPL-PO expects Miami to be caught up and begin generating NRT AVHRR SST data in early CY 1998. Once distribution of the NRT data begins, SDST will be able to work the specifics of how to acquire it and make it available on the larry server for use ultimately by ECS and the MOD27Y science code. Sample data sets are available now from the JPL-PO Web site (http://podaac.jpl.nasa.gov/sst/). At present, there are no plans to offer this data set in HDF-EOS; only the native HDF format will be available at the GSFC DAAC.

Frequency (files/day) = TBD

File size (MB/file) = TBD

Volume (MB/day) = TBD

Native Format = HDF

Format Ultimately Available for V2: Native HDF

4.3.5 TOMS Daily Column Ozone

Interest in this data comes from the University of Wisconsin and the Tanre atmosphere groups. The data comes from the new Earth Probe and ADEOS satellites, which were launched in the summer of CY 1996. Both satellites have been turned on and are transmitting raw data. TOMS Column Ozone NRT will be available at the GSFC DAAC in both native and HDF formats (**NOT** HDF-EOS!). ECS believes (based on comments in the ECS paper "Pre-processing of NESDIS and TOMS Ancillary Data Sets, by Shaun de Witt, dated 4/96, document No. 170-TP-002-001) that there is insufficient "value-added" in pre-processing the TOMS data into HDF-EOS to warrant producing an HDF-EOS version of TOMS data sets.

NRT daily ozone data is being generated from the raw data by the Atmospheric Chemistry Group at GSFC and is available on the "jwocky" server (ftp: jwocky.gsfc.nasa.gov). The product files are fetched once a day by the larry server and made available via an anonymous ftp site.

jwocky Earth Probe pathname: /pub/eptoms/ozYY/gaYYMMDD.ept

larry pathname: /pub/eptoms/ozYYMM/gaYYMMDD.ept

Frequency (files/day) = TBD

File size (MB/file) = TBD

Volume (MB/day) = TBD

Native Format = ASCII

Formats Ultimately Available for V2: ASCII; HDF

jwocky ADEOS pathname: /pub/adeos/ozYY/gaYYMMDD.alt

larry pathname: /pub/adeos/ozYYMM/gaYYMMDD.alt

Frequency (files/day) = TBD

File size (MB/file) = TBD

Volume (MB/day) = TBD

Native Format = ASCII

Formats Ultimately Available for V2: Native; HDF

4.3.6 1-Degree NCEP GDAS

SeaWiFS (and, therefore by implication, the Miami oceans group and MODIS) will change to this data set for their Met model input by the middle of April 1997; Wisconsin will change to this data set for their Met model input too, but not until after launch.

As with the 2.5 degree GDAS data set, the 1 degree GDAS data set is distributed by the NCEP nic server, fetched daily by the larry server, and made available via an anonymous ftp site.

nic pathname: pub/fnl/gdas1.TXXZ.PGrbF00 (XX = 00, 06, 12, 18)

larry pathname:
/pub/ncep_data/anal_YYMM/anal_YYMMDD/gdas1.PgrbF00.YYMMDD.HHZ
(HH = 00, 06, 12, 18)

Frequency (files/day) = 4

File size (MB/file) = 1.97

Volume (MB/day) = 4 x 1.97 = 7. 88

Native Format = Grib

Formats Ultimately Available for V2: Grib, HDF-EOS

4.3.7 MISR

Both the Montana and the Boston Land teams will eventually be using data from this AM-1 instrument. Wolfgang Wanner of the Boston team is working directly with the MISR team. The Montana products will eventually incorporate MISR data too.

Up to now, MISR has been considered ancillary data for MODIS. However, in SDST's view MISR should be redefined out of that category, as the AM-1 scientists involved

are working directly with each other, and the DAAC will have no interface to a non-EOS data source. It is thus not clear what SDST's contribution to this effort should be other than to offer their assistance to help facilitate interaction between MISR Science team and the interested MODIS Land teams.

4.3.8 Buoy Time-and-Location Data

Bob Evans, University of Miami Remote Sensing Group, will periodically send this data to the GSFC DAAC (usually daily, and yearly for a special summary data set). The data will be fetched by the GSFC DAAC's Data Ingestion Sub-system and be used as input to one or more PGEs. These PGEs will extract Level-1A, Level-1B, and Level-2 MODIS product data and ancillary data at and near the buoys. This extracted data will then be sent back to the Miami where it will be used as input to their ocean color and SST data matchup process.

Nothing has been determined yet about the details of this back-and-forth data transfer process or the native format of the data itself. A DAAC server will probably have to go at a-periodic intervals (intervals still TBD) to fetch the buoy time-and-location file. So, although it is an intra-MODIS data transfer, it looks functionally very much like ancillary data and it makes sense to treat it that way. The specification of the a-periodic intervals is dependent somewhat on how long MODIS data can remain staged and how resource intensive it is to restage data.

Working out the data transfer and processing details will require some close work between Miami, SDST, and ECS. Only the native format will be available at the GSFC DAAC.

Frequency (files/day) = 1 (plus the yearly summary file)

File size (MB/file) = TBD

Volume (MB/day) = TBD

Native Format = ASCII

Format Ultimately Available for V2: Native

4.3.9 AIRS/AMSU

This is a post-launch need, for Wisconsin. There is no obvious need to do anything about this yet since AIRS/AMSU is a PM-1 instrument. Since it is an EOS project, its processing will presumably occur at the DAAC. SDST suggests that the MISR model applies here, in that the most logical way for this to be addressed may be for the Wisconsin and AIRS/AMSU science teams to work directly together, and to essentially drop this from the "ancillary data" domain.

4.3.10 NOAA NRT ADEOS NSCAT Overwater Surface Winds

Requested by the Tanre Atmosphere group. This data comes from the new ADEOS satellite, which was launched in August of 1996. The satellite has been turned on and is transmitting raw data.

Experimental NRT over water surface wind data is being generated from the raw data by a NOAA/NESDIS Group co-located with NCEP in Suitland, Maryland and the JPL NSCAT Processing Group. It is not known yet whether this data set will be acquired from the JPL NSCAT Processing Group or from NOAA/NESDIS, Both groups will ultimately be producing somewhat different versions of the NRT data set. The native format of both versions of the data set are TBD. At present, there are no plans to offer this data set in HDF-EOS; only the native format will be available at the GSFC DAAC.

Frequency (files/day) = TBD

File size (MB/file) = TBD

Volume (MB/day) = TBD

Native Format = TBD

Formats Ultimately Available for V2: Native

5. ALTERNATE DATA SOURCES

In the production environment, there will be times when ancillary data inputs will not be available. It is desirable that the science team have some plans to deal with these situations. One approach is to consider the use of substitute (alternate) data sets. So far there are no formal requirements for alternates to be specified. Since some MODIS products may be backed up by ancillary data sets, and vice versa, thinking about alternate data sets quickly takes one beyond ancillary data per se, and into the general question of production rules.

There has been some unfocused discussion of production rules as they relate to the availability of input data sets. Please see Section 6 for a discussion of this problem and a strawman proposal as to how it might be approached.

Given that the MODIS approach to missing input data sets has yet to be worked out in detail, the following is SDST's recommendation as to how science team members should approach the problem with a reasonable degree of assurance that their approach will be satisfactory no matter what the ultimate production rules turn out to be as regards missing inputs.

In SDST's view, the best way for an SCF to address the problem of missing inputs is to make some decisions about a hierarchy of backups, code accordingly, and tell SDST what data sets the SCF wants available. In that sense SCFs can erase the formal distinction between primary and secondary ancillary data sets except as that distinction is embodied in if-blocks in their code. This approach should stand SCFs in good stead, if it is taken in the spirit of doing everything one possibly can to keep a process running and producing scientifically useful data despite the possibility of missing inputs. (See Section 6 for some elaboration of the philosophy behind this approach.)

6. OUTSTANDING ISSUES

6.1 Missing Input Data Sets

One of the major unresolved issues for MODIS as we go from Version 1 into Version 2 is how to handle the problem of missing inputs to MODIS processes, either MODIS products which are not available when needed as inputs, or ancillary data sets which are not available when needed.

Some MODIS products and ancillary data sets can be considered as possible alternates for each other when the desired data set is not available. However, both the use and the non-use of alternates raise a number of issues which go beyond the domain of ancillary data per se. The purpose of this proposal is to highlight what SDST thinks some of the questions are by putting up a suggested MODIS approach to this issue.

6.1.1 Operational Problem: Lack of a Specified Input

SDST sees three basic responses to a runtime lack of a specified input:

Response 1: Stop processing until the input is available again. This would occur for cases where the input is critical to the output and there is no acceptable substitute.

Response 2: Do without the input and keep processing. This would occur where the input is desirable but not critical and there is either:

(a) no acceptable substitute, or

(b) available substitutes do not confer sufficient advantage to be worth bothering with.

Response 3: Use a substitute input and keep processing. This is for cases in which substitutes are acceptable, available, and confer enough advantage over doing without so as to be worth bothering with.

In Response 3, the input is either a MODIS product or an ancillary data stream, and the substitute could be either a MODIS product, an ancillary data stream, or a static ancillary data set (such as a climatology).

The approach that SDST is advocating is to make it a MODIS goal to avoid Response 1 whenever possible, but to leave the choice of when in fact to invoke Response 1 up to the individual SCF team producing the product. Similarly, when Response 1 is not chosen, the choice between Response 2 and Response 3 would be up to the SCF.

The only formal requirement would be that for each product, the responsible team formally specify under what conditions of missing inputs it would be necessary to stop its processing. This formal specification would allow SDST to develop a tree of "go/no-go" dependencies for MODIS processing.

Two questions that occur under the suggested scenario are:

1. When do we decide that a late input product is to be considered as not available?
2. How long do we wait if we don't know what the status of an input product is?

6.1.2 Role of SDST

If alternate ancillary data sources are to be turned to, they have to be readable by the executing process' code, like any other ancillary data source. All that is needed under the proposed scenario is the normal ancillary data activity, in which the particular SCF team works with SDST to get the necessary files identified and added to the ancillary data requirements given to ECS.

6.1.3 "Consumer Disclosure" for End Users of MODIS Products

Changes in the inputs to a product are something that users of a product might well want to know about. Inputs (ancillary and otherwise) should be listed in the ECS Inventory metadata (in the INPUTPOINTER attribute) for each MODIS product file. For example, if a particular product used MODIS ozone to make one output file and TOMS ozone to make another, then that difference should be reflected in the ECS metadata for each of those files. Sufficiently curious end users could then trace the chain of inputs as far back as they wish to do so, but end-user concerns in regard to consistency of inputs would not dictate when products are produced.

Additional MODIS product input information can be included in an ECS Inventory or Product Specific Attribute (PSA) Archive Metadata if this information is deemed important. If a search capability for the additional MODIS product input information is necessary, ECS Inventory metadata should be used to store the PSA. If a search capability for the additional MODIS product input information is not deemed important, ECS archive metadata could be used to store the PSA product input information. Unlike ECS Inventory metadata, the ECS Archive metadata is not searchable.

A question that needs to be resolved when this approach of "caveat emptor" metadata is implemented is, "How far down toward the root of the processing tree should a product's inputs be identified?" One level down would be sufficient to enable a motivated user unambiguously to trace the input history all the way back to the radiances, but would a more informative approach be better? Or would it lead to an unmanageable clutter of input tracing? Would it be desirable to have a yes-no flag in the metadata which indicates whether all the normal inputs were present?

6.2 Version 2 Ancillary Data Issues to be Resolved

There are a number of ancillary data related issues that still remain to be resolved for MODIS V2. These are presently being worked by SDST. They are summarized and informally discussed in Appendix B.

(This page intentionally left blank.)

APPENDIX A: SUMMARY OF TIME-VARYING V2 ANCILLARY DATA

Data Set Name	Data Set Source Server	Available Formats in DAAC	Requester						
			Atmos.-Univ. of Wisc	Atmos.-Tanre	Ocean-Univ. of Miami	Ocean-Esaias & Turpie	Land-GSFC	Land - Boston	Land - Montana
1 Degree NCEP AVN Analysis	NCEP nic	Grib and HDF-EOS	X						
1 Degree NCEP GDAS	NCEP nic	Grib and HDF-EOS	X (PL)		X				
DAO High Time Resolution Global Analysis Output for AM-1	DAO (TBD)	HDF-EOS			X		X		X
TOVS Column Ozone Twice-Daily Product	NCEP nic	Grib and HDF-EOS			X				
TOVS Column Ozone Daily Product	NCEP nic	Grib and HDF-EOS	X						
TOMS Column Ozone (ADEOS)	GSFC Atmosphere Chemistry Group jwocky	ASCII and HDF	X						
TOMS Column Ozone (Earth Probe)	GSFC Atmosphere Chemistry Group jwocky	ASCII and HDF	X						
Reynolds Weekly SST	NCEP nic Server	ASCII and HDF-EOS	X						
Pathfinder AVHRR Daily SST	JPL Phys. Ocean DAAC (TBD)	HDF				X			
SSM/I Sea Ice	NCEP polar	Grib and HDF-EOS	X						
NOAA NRT ADEOS NSCAT Overwater Surface Wind	NOAA/NESDIS or JPL (TBD)	Native TBD		X					
FNMOC Global Mixed-Layer Model Output	WHOI or Commercial Provider (TBD)	HDF				X			
Buoy Time/Location Data	University of Miami (TBD)	ASCII			X				
AIRS/AMSU (MODIS PL Ancillary Data– Included for Completeness)	PM-1		X (PL)						
MISR (Not Considered Ancillary Data since MISR is AM-1 Instrument-Like –Included for Completeness)	AM-1							X	X
Note: University of Miami = Same as SeaWiFS									

(This page intentionally left blank.)

APPENDIX B: ANCILLARY DATA OUTSTANDING ISSUES

The issues informally presented and discussed below are actively being worked by SDST. As answers become available, they will be transmitted to interested and affected members of the MODIS Science Team and SDST.

B.1 Science Team Issues

TOMS Data sets

1. Is the Native TOMS format the same for all satellite platforms?
   A. If the answer to is no, then should some sort of pre-processor be developed to convert the various native TOMS formats to some common format that minimizes the differences between the native formats?

2. The DAAC will be converting all TOMS data to HDF (**NOT** HDF-EOS). If there are different native TOMS formats for different platforms, should we recommend to the MODIS Science Team that they use the HDF version of TOMS to simplify their coding for TOMS Data?

NSCAT Data set

3. The NSCAT NRT data needed by the Atmosphere Science Team will probably have to be spatially and temporally averaged (i.e. a L3 product will have to be produced from the available L2 product granules) to fulfill the Team's needs.
   A. Who will do this? (The Atmosphere Science Team could do this using the Oceans L3 space and Time Bin Routines or the Atmospheres V2 tiling routines for L3 if necessary but if that is done, it should be a cooperative venture between SDST and the Atmosphere Science Team).

DAO Data sets

4. Which variables/parameters will be required by the prospective users (Land and Oceans Science Teams) of the high time resolution DAO data set? This question has been raised by DAO in an attempt to see if the desired data is a sufficiently small data set that it may be worthwhile for them to rethink their reluctance to subset the data-stream before sending operational high time resolution data to the DAAC.

   A. What data sets will be required for V2?

   B. What additional data sets will be required PL?

B.2 Data Source Issues

AVHRR Pathfinder SST Data set

1. How realistic is the estimated time (early CY 1998) given by JPL for when Miami will switch over from retrospective processing to NRT Processing of the Pathfinder SSTs?

2. When NRT processing begins, will the JPL DAAC support ftp implemented standing data orders for the NRT SST files that MODIS needs? Alternatively, will the larry server have to use a CRON job to fetch (by ftp) the Pathfinder SST data from the JPL DAAC?

Reynolds SST Data

4. This data set is a weekly average SST product. The date assigned to a product file is the date of the Wednesday of the week over which the average is taken. It appears that the week associated with a Reynolds SST file runs from Sunday to Saturday inclusive or from Monday to Sunday inclusive. Which of these two "weeks" is the time period actually covered by a Reynolds data set file?

B.3 Production System Issues

Data sets which are in Native Format

1. How will the DAAC physically make available the ancillary data sets in native format?
   A. Will they be available in some (as yet) specified sub-directory?

The Navy FNMOC Ocean Mixed Layer Model Output Data Set

2. How will the FNMOC Ocean Mixed Layer Depth be accessed by the DAAC?

   A. Can it be fetched by the DAAC's larry server from an appropriate WHOI server or must it be fetched from the Navy's designated commercial provider?

   B. If the answer to 2.A is the Mixed Layer Depth product can be fetch from an appropriate WHOI server:

      1) Which WHOI server will it be placed on?

      2) Who has the responsibility at WHOI and ESDIS to negotiate the terms of the agreements and reports [Memorandum of Understanding (MOU), Interface Requirements Document (IRD), and Interface Control Document (ICD)] that

document how ECS/ESDIS will acquire the FNMOC Mixed Layer model output files from the WHOI?

C. If the answer to 2.A is the Mixed Layer Depth product must be fetch from the commercial provider server:

1) Can the commercial provider of the FNMOC fields (including the Mixed Layer Depth) modify his/her setup to include Internet ftp access as required by ECS/ESDIS?

2) Will he/she modify the setup to include Internet ftp access as required by ECS/ESDIS?

3) Will the commercial provider charge ECS/ESDIS for the modification?

4) Who has the responsibility at WHOI and ESDIS to negotiate the terms of the agreements and reports (MOU, IRD, and ICD) that document how ECS/ESDIS will acquire the FNMOC Mixed Layer model output files from the WHOI?

5) How long will it take for ECS/ESDIS to write draft copies of the MOU, IRD, and ICD addressing the acquisition of the FNMOC Mixed Layer Depth data by ECS/ESDIS.

6) When will Matt Schwaller's ECS/ESDIS Ancillary Data Group rewrite and resubmit the Configuration Change Requests (CCRs) that has previously been rejected by the ESDIS Configuration Control Board (CCB) to acquire the FNMOC Mixed Layer Depth data.

7) Once the agreements, reports, and ESDIS CCR have been approved by all parties, how long will it take to get the appropriate scripts written, tested, and made operational to fetch the Mixed Layer product from WHOI to the larry server?

Backup Ancillary Data

3. Have appropriate backup data sets been identified by the MODIS Atmosphere, Ocean, and Land Science Teams for V2?

A. What is the PGE Production Rule syntax for specifying Backup Ancillary Data Files?

B.4 ECS Issues

Administrative Issues

1. There does not appear to be any provision yet by ECS to develop scripts that automatically query the DAAC larry server and NOAA data server(s) for the appearance of the next available ancillary file that will be ingested into the Ingest

Server. No scripts appear to have been developed either to automatically bring back data sets that are needed once they are available.

A. Are these scripts going to be written by ECS?

B. When will they be written?

C. When will they become available?

D. Are there any documents that lay out requirements and design for these scripts?

1) If there are, what documents are they?

2. ESDIS CCRs must be prepared by MODIS to acquire or document ancillary data needed by MODIS science software when run at the GSFC or EDC DAAC. They are submitted to the ESDIS CCB through Matt Schwaller of ESDIS. Will MODIS be allowed to send a representative to attend the meeting(s) in which the ESDIS CCRs are scheduled to be acted upon?

Data sets which are in Native Format

3. Since all files need metadata in the ECS system, when will ECS develop the software to generate the separate metadata file that must accompany each ancillary file when it is in native format?

A. Does ECS have requirements and a design for this software?

B. What document(s) contain the requirements and design for this software?